

AN ASYMPTOTIC TOTAL VARIATION TEST FOR COPULAS

JEAN-DAVID FERMANIAN, DRAGAN RADULOVIĆ, AND MARTEN H. WEGKAMP

ABSTRACT. We propose a new goodness-of-fit test for copulas, based on empirical copula processes and their nonparametric bootstrap counterparts. The standard Kolmogorov-Smirnov type test for copulas that takes the supremum of the empirical copula process indexed by half spaces is extended by test statistics based on the supremum of the empirical copula process indexed by families of L_n disjoint rectangles, with L_n slowly tending to infinity. Although the underlying empirical process does not converge, it is proved that the p-values of our new test statistic can be consistently estimated by the bootstrap. Simulations confirm that the power of the new procedure is often higher than the power of the standard Kolmogorov-Smirnov or the Camèr-von Mises tests for copulas.

RUNNING TITLE: An Asymptotic Total Variation Test for Copulas

MSC2000 SUBJECT CLASSIFICATION: Primary 60F17 ; secondary 60G99.

KEYWORDS AND PHRASES: Bootstrap, Copula, Empirical Copula Process, Goodness-of-Fit Test, Weak Convergence.

1. INTRODUCTION

This paper introduces a new powerful goodness-of-fit (GOF) test for bivariate copulas in $[0, 1]^2$ based on the empirical copula process

$$(1) \quad \mathbb{Z}_n(u, v) = \sqrt{n}(\mathbb{C}_n - C)(u, v), \quad 0 \leq u, v \leq 1,$$

given a sample of n independent observations (X_i, Y_i) ¹. Here $\mathbb{C}_n = \mathbb{H}_n(\mathbb{F}_n^-, \mathbb{G}_n^-)$ is the empirical copula, introduced by Deheuvels (1979), \mathbb{H}_n is the joint cdf of the sample $(X_1, Y_1), \dots, (X_n, Y_n)$ and \mathbb{F}_n^- and \mathbb{G}_n^- denote the marginal quantile functions of the sample X_1, \dots, X_n and Y_1, \dots, Y_n , respectively. The Kolmogorov-Smirnov (KS) test

The research of Wegkamp is supported in part by NSF Grant DMS-10-07444.

¹ For notational convenience, we restrict ourselves to the two-dimensional case. All our results extend to copulas in $[0, 1]^d$ with $d > 2$ fixed.

statistic for testing of the null hypothesis $H_0 : C = C_0$ is

$$(2) \quad \text{KS}_n = \sup_{0 \leq u, v \leq 1} |\sqrt{n}(\mathbb{C}_n - C_0)(u, v)|.$$

The Cramèr-von Mises statistic (CvM) is

$$(3) \quad \text{CM}_n = \int \{\sqrt{n}(\mathbb{C}_n - C_0)(u, v)\}^2 d\mathbb{C}_n(u, v).$$

It is well-known, see, for instance, Fermanian et al. (2004), that \mathbb{Z}_n and its bootstrap counterpart \mathbb{Z}_n^* , defined in (14) below, both converge weakly to the same tight Gaussian process in $\ell^\infty([0, 1]^2)$ under the null hypothesis. Therefore, we can compute the α -upper points of KS_n and CM_n via the bootstrap. To the best of our knowledge, all the proposed GOF tests rely on simulation-based procedures to calculate their corresponding p-values². A parametric bootstrap has been proposed (Genest and Rémillard, 2008) to tackle composite null hypotheses, while Rémillard and Scaillet (2009) advocate the use of the multiplier central limit theorem to build an alternative bootstrap empirical copula process. Bücher and Dette (2010) gives a survey and a comparison of various bootstrap methods.

The goal of this paper is to develop a more powerful test than the KS test (2) and CvM test (3). We propose the following test that rejects the null hypothesis $H_0 : C = C_0$ for large values of the test statistic

$$(4) \quad \mathbb{T}_n := \sup_{B_1, \dots, B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Z}_n(B_k)|.$$

The supremum is taken over all *disjoint* rectangles $B_1, \dots, B_{L_n} \subset [0, 1]^2$ of the form $(a, b] \times (c, d]$, using the convention

$$(5) \quad \mathbb{Z}_n((a, b] \times (c, d]) = \mathbb{Z}_n(b, d) - \mathbb{Z}_n(a, d) - \mathbb{Z}_n(b, c) + \mathbb{Z}_n(a, c),$$

for all $0 \leq a \leq b \leq 1$ and $0 \leq c \leq d \leq 1$.

²with the notable exception of the distribution-free test statistics of Fermanian (2004). This idea has been further developed by Scaillet (2007) and Fermanian and Wegkamp (2012).

Now, if $L_n = L$ for all n , the collection of rectangles is sufficiently small that we can still appeal to the weak convergence of \mathbb{Z}_n and \mathbb{Z}_n^* in conjunction with the continuous mapping theorem, to obtain α -upper points of the test statistic \mathbb{T}_n via the bootstrap. Taking $L_n = +\infty$ for all n , or equivalently, if we consider all families of disjoint rectangles in $[0, 1]^2$ (possibly partitions), the statistic \mathbb{T}_n is equal to the total variation distance $TV(\mathbb{Z}_n)$ of \mathbb{Z}_n . The resulting test isn't statistically meaningful as $TV(\mathbb{Z}_n)$ is maximal, to wit, $TV(\mathbb{Z}_n) = n^{1/2} \rightarrow +\infty$. The problem is to find a rich collection that quickly detects departure from the null, but still yields a consistent test. The main novelty of our approach is the fact that we let L_n , the number of rectangles, slowly tend to ∞ in that $L_n \sim (\log n)^\gamma$, $0 < \gamma < 1$. While in this case the process \mathbb{Z}_n no longer converges, Theorem 1 in Section 2 states that we can still consistently estimate the distribution of the process \mathbb{Z}_n by the bootstrap. We refer to our procedure as the Asymptotic Total Variation (ATV) test. The considered families of rectangles are finer and finer, presumably improving the power of the test, while for each n large enough, we still have a consistent test in that we control the type 1 error. A key observation is that under the null hypothesis $H_0 : C = C_0$, we have $\mathbb{T}_n \leq L_n \sup_B |\mathbb{Z}_n(B)| = O_p(L_n)$, while under the alternative $H_A : C = C_1$ for some fixed $C_1 \neq C_0$, \mathbb{T}_n is much larger since the bias is at least of order $O(n^{1/2})$.

Theorem 1 extends the surprising result obtained by Radulović (2012) for empirical processes indexed by sums of indicator functions of VC-graph classes (Theorem 12 in the appendix). We require very mild conditions on the copula function C . This is one of the few notable exceptions known to us in the literature where the bootstrap “works”, that is, the conditional bootstrap distribution consistently estimates the distribution of the test statistic, while the distribution of the statistic itself does not converge. For other instances of this phenomenon, we refer to Bickel and Freedman (1983) or Radulović (1998, 2012), more recently.

Section 3 considers the more general hypothesis that the underlying copula C belongs to some parametric copula family $\{C_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$. Given a sufficiently regular

estimator $\hat{\theta}$ and its bootstrap counterpart $\hat{\theta}^*$, we adjust our statistic (4) and its non-parametric bootstrap counterpart to obtain a consistent level α test (Theorem 3). Again, the result is established under very mild regularity conditions on the copula C_θ and the estimators $\hat{\theta}$ and $\hat{\theta}^*$.

Section 4 then reports a small numerical study where we show that, in complex but realistic situations, our test (4) is superior to the Kolmogorov-Smirnov and the Cramèr-von Mises tests. We also comment on a possible inadequacy in the way the copula GOF tests are commonly evaluated.

Finally, the proofs are collected in Section 5. The appendix contains some technical results from Segers (2012) and Radulović (2012).

2. GENERALIZED KOLMOGOROV-SMIRNOV TEST

NOTATION. Let H be the distribution function of the random variables (X, Y) with marginals F and G . We will assume throughout the paper that H , F and G are continuous. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent copies of (X, Y) . We denote the generalized inverses of F and G by F^- and G^- , respectively. For instance,

$$(6) \quad F^-(u) = \inf\{x \mid F(x) \geq u\}.$$

The empirical counterparts of H , F and G are, respectively,

$$(7) \quad \begin{aligned} \mathbb{H}_n(x, y) &= \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x, Y_i \leq y\}, \\ \mathbb{F}_n(x) &= \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq x\}, \\ \mathbb{G}_n(y) &= \frac{1}{n} \sum_{i=1}^n 1\{Y_i \leq y\}, \end{aligned}$$

with $-\infty < x, y < \infty$. We denote the copula function of (X, Y) by

$$(8) \quad C(u, v) = H(F^-(u), G^-(v)), \quad 0 \leq u, v \leq 1$$

and its empirical estimate by

$$(9) \quad \mathbb{C}_n(u, v) = \mathbb{H}_n(\mathbb{F}_n^-(u), \mathbb{G}_n^-(v)), \quad 0 \leq u, v \leq 1.$$

The empirical copula process $\mathbb{Z}_n(u, v) = \sqrt{n}(\mathbb{C}_n - C)(u, v)$ is already defined in (1). We define \mathcal{F}_n as the class of functions

$$(10) \quad f(x, y) = \sum_{k=1}^{L_n} c_k 1\{(x, y) \in B_k\},$$

with $c_k \in \{-1, +1\}$ and *disjoint* rectangles B_k of the form $(a, b] \times (c, d]$ in the unit square $[0, 1]^2$, for all $1 \leq k \leq L_n$. We let

$$\mathbb{Z}_n(f) = \sum_{k=1}^{L_n} c_k \mathbb{Z}_n(B_k),$$

and observe that

$$\mathbb{T}_n = \sup_{f \in \mathcal{F}_n} |\mathbb{Z}_n(f)| = \sup_{B_1, \dots, B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Z}_n(B_k)|,$$

where the supremum is taken over all disjoint rectangles B_1, \dots, B_{L_n} of the unit square $[0, 1]^2$.

If $L_n = L$ for all n , then $\mathcal{F}_n = \mathcal{F}$ and \mathbb{Z}_n converges in $\ell^\infty(\mathcal{F})$ to a Gaussian process under regularity conditions on C , see, for instance, Fermanian et al. (2004) and Segers (2012). As a consequence of the continuous mapping theorem, \mathbb{T}_n trivially converges weakly as well. However, if $L_n \rightarrow \infty$, as $n \rightarrow \infty$, this is no longer true as the process \mathbb{Z}_n does not converge weakly.

The main point of this paper is to show that, provided $L_n = (\log n)^\gamma$ for some $0 < \gamma < 1$, the distribution of \mathbb{T}_n can be estimated by the bootstrap. The bootstrap counterparts of the above processes are defined as follows. Let the bootstrap sample $(X_1^*, Y_n^*), \dots, (X_n^*, Y_n^*)$ be obtained by sampling with replacement from $(X_1, Y_1), \dots, (X_n, Y_n)$. We write

$$(11) \quad \mathbb{H}_n^*(x, y) = \frac{1}{n} \sum_{i=1}^n 1\{X_i^* \leq x, Y_i^* \leq y\}$$

for the empirical cdf based on the bootstrap, with marginals

$$(12) \quad \mathbb{F}_n^*(x) = \frac{1}{n} \sum_{i=1}^n 1\{X_i^* \leq x\}$$

and

$$(13) \quad \mathbb{G}_n^*(y) = \frac{1}{n} \sum_{j=1}^n 1\{Y_j^* \leq y\},$$

and we denote its associated empirical copula function by \mathbb{C}_n^* . The bootstrap empirical copula process is

$$(14) \quad \mathbb{Z}_n^* = \sqrt{n}(\mathbb{C}_n^* - \mathbb{C}_n) = \sqrt{n} \left\{ \mathbb{H}_n^*(\mathbb{F}_n^{*-}, \mathbb{G}_n^{*-}) - \mathbb{H}_n(\mathbb{F}_n^-, \mathbb{G}_n^-) \right\}.$$

ASSUMPTIONS. We will assume the following set of assumptions:

- (C1) The first-order partial derivatives $C_1(u, v) = \partial C(u, v)/\partial u$ and $C_2(u, v) = \partial C(u, v)/\partial v$ of $C(u, v)$ exist, are of bounded variation, and satisfy, for some $r > 0$, $\beta \geq 0$ and $K < \infty$,

$$|C_1(u, v) - C_1(s, t)| \leq K \left(u^{-\beta}(1-u)^{-\beta} + s^{-\beta}(1-s)^{-\beta} \right) (|u-s|^r + |v-t|^r)$$

and

$$|C_2(u, v) - C_2(s, t)| \leq K \left(v^{-\beta}(1-v)^{-\beta} + t^{-\beta}(1-t)^{-\beta} \right) (|u-s|^r + |v-t|^r)$$

for all $0 < u, v, s, t < 1$.

- (C2) The number L_n is of order $(\log n)^\gamma$ for some $0 < \gamma < 1$.

REMARK. We know that continuity of the partial derivatives of $C(u, v)$ on $(0, 1)^2$ is required for weak convergence, see Fermanian et al. (2004) and Segers (2012). The requirement that the partial derivatives are of bounded variation is natural since we compute the supremum of \mathbb{Z}_n over increasingly finer families of rectangles in $[0, 1]^2$, and the process $\mathbb{Z}_n(u, v)$ is asymptotically equivalent to $\alpha_n(u, v) - C_1(u, v)\alpha_n(u, 1) - C_2(u, v)\alpha_n(1, v)$ with $\alpha_n(u, v) = \sqrt{n}(\mathbb{H}_n(u, v) - H(u, v))$, see Proposition 9. The additional requirement is weaker than imposing a Hölder condition on the derivatives. Segers (2012) imposes a slightly stronger condition on the second partial derivatives

of C (corresponding to $r = 1$) to obtain an almost sure representation of the empirical copula process. The second assumption (C2) allows for sub-logarithmic rate in the sample size for the number of rectangles considered. In practice, even this fairly slow rate yields much better tests, see our simulations in Section 4.

Our first result states that the processes \mathbb{Z}_n and \mathbb{Z}_n^* are close in the bounded Lipschitz distance d_{BL_1} that metrizes weak convergence. Formally, we define

$$(15) \quad d_{BL_1}(\mathbb{Z}_n, \mathbb{Z}_n^*) = \sup_h |\mathbb{E}[h(\mathbb{Z}_n)] - \mathbb{E}^*[h(\mathbb{Z}_n^*)]|.$$

Here \mathbb{E}^* is the conditional expectation with respect to the bootstrap sample. The supremum in (15) is taken over the class $BL_1 := BL_1(\ell^\infty(\mathcal{F}_n))$ of all functionals $h : \ell^\infty(\mathcal{F}_n) \rightarrow \mathbb{R}$ with

$$(16) \quad \sup_{x \in \ell^\infty(\mathcal{F}_n)} |h(x)| \leq 1$$

and

$$(17) \quad |h(x) - h(y)| \leq \sup_{f \in \mathcal{F}_n} |x(f) - y(f)|$$

for all $x, y \in \ell^\infty(\mathcal{F}_n)$.

THEOREM 1. *Let $\mathbb{Z}_n = \{\mathbb{Z}_n(f), f \in \mathcal{F}_n\}$ and $\mathbb{Z}_n^* = \{\mathbb{Z}_n^*(f), f \in \mathcal{F}_n\}$ with \mathcal{F}_n as defined in (10) above. Under conditions (C1) and (C2), we have*

$$(18) \quad d_{BL_1}(\mathbb{Z}_n, \mathbb{Z}_n^*) = \sup_{h \in BL_1} |\mathbb{E}[h(\mathbb{Z}_n)] - \mathbb{E}^*[h(\mathbb{Z}_n^*)]| \rightarrow 0,$$

in probability, as $n \rightarrow \infty$.

Since $d_{BL_1}(\mathbb{Z}_n, \mathbb{Z}_n^*) \leq 2$, an equivalent statement is that

$$\lim_{n \rightarrow \infty} \mathbb{E}[d_{BL_1}(\mathbb{Z}_n, \mathbb{Z}_n^*)] = 0.$$

Since the mapping $h(X) = \sup_{f \in \mathcal{F}_n} |X(f)|$ belongs to BL_1 , Theorem 1 implies that, for any Lipschitz function g ,

$$(19) \quad \left| \mathbb{E} \left[g \left(\sup_{f \in \mathcal{F}_n} |\mathbb{Z}_n(f)| \right) \right] - \mathbb{E}^* \left[g \left(\sup_{f \in \mathcal{F}_n} |\mathbb{Z}_n^*(f)| \right) \right] \right| \rightarrow 0,$$

in probability, as $n \rightarrow \infty$. The idea is to approximate the distribution of the statistic \mathbb{T}_n by the conditional (bootstrap) distribution of

$$(20) \quad \mathbb{T}_n^* = \sup_{f \in \mathcal{F}_n} |\mathbb{Z}_n^*(f)| = \sup_{B_1, \dots, B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Z}_n^*(B_k)|.$$

COROLLARY 2. *Consider the null hypothesis $H_0 : C = C_0$, for some copula function C_0 satisfying (C1). Then the test statistic*

$$(21) \quad \mathbb{T}_n = \sqrt{n} \sup_{f \in \mathcal{F}_n} |\mathbb{C}_n(f) - C_0(f)|$$

with \mathcal{F}_n as described in (10) above, and with L_n satisfying (C2), can be consistently estimated by \mathbb{T}_n^ , under the null hypothesis $H_0 : C = C_0$. That is,*

$$(22) \quad \sup_{h \in BL_1} |\mathbb{E}[h(\mathbb{T}_n)] - \mathbb{E}^*[h(\mathbb{T}_n^*)]| \rightarrow 0,$$

in probability, as $n \rightarrow \infty$.

REMARK. Note that, in particular, these results continue to apply if we replace \mathcal{F}_n by a subset $\tilde{\mathcal{F}}_n \subset \mathcal{F}_n$. For instance, we could restrict ourselves on regular partitions (based on grids in $[0, 1]^2$). This may be important to calculate \mathbb{T}_n in practice.

REMARK. The Corollary states that $d_{BL_1}(\mathbb{T}_n, \mathbb{T}_n^*) \rightarrow 0$ in probability, as $n \rightarrow \infty$. For our test, we need the α -upper point of the statistic \mathbb{T}_n , that we approximate by the bootstrap \mathbb{T}_n^* . The bootstrap quantile approximation is justified by the following simple observation. Let $\varepsilon > 0$ be arbitrary and define the Lipschitz function

$$g_{t,\varepsilon}(x) = 1\{x \leq t\} + \frac{t + \varepsilon - x}{\varepsilon} 1\{t < x \leq t + \varepsilon\}.$$

We have, for $\delta_n := d_{BL_1}(\mathbb{T}_n, \mathbb{T}_n^*)$,

$$\begin{aligned} \mathbb{P}\{\mathbb{T}_n \leq t\} &\leq \mathbb{E}[g_{t,\varepsilon}(\mathbb{T}_n)] \\ &= \mathbb{E}^*[g_{t,\varepsilon}(\mathbb{T}_n^*)] + \mathbb{E}[g_{t,\varepsilon}(\mathbb{T}_n)] - \mathbb{E}^*[g_{t,\varepsilon}(\mathbb{T}_n^*)] \\ &\leq \mathbb{P}^*\{\mathbb{T}_n^* \leq t + \varepsilon\} + \delta_n/\varepsilon, \end{aligned}$$

since $g_{t,\varepsilon}$ has Lipschitz constant $1/\varepsilon$. A similar computation shows that

$$\mathbb{P}^*\{\mathbb{T}_n^* \leq t - \varepsilon\} - \delta_n/\varepsilon \leq \mathbb{P}\{\mathbb{T}_n \leq t\}$$

so that, uniformly in t ,

$$(23) \quad \mathbb{P}^* \{ \mathbb{T}_n^* \leq t - \varepsilon \} - \delta_n / \varepsilon \leq \mathbb{P} \{ \mathbb{T}_n \leq t \} \leq \mathbb{P}^* \{ \mathbb{T}_n^* \leq t + \varepsilon \} + \delta_n / \varepsilon$$

Since \mathbb{T}_n^* is a discrete random variable with finitely many atoms, we have

$$\mathbb{P}^* \{ \mathbb{T}_n^* \leq t - \varepsilon \} = \mathbb{P}^* \{ \mathbb{T}_n^* \leq t + \varepsilon \}$$

for almost all $t \in \mathbb{R}$, and for $\varepsilon > 0$ small enough. This equality can be verified in practice for any given t . Then, for such a t , since $\delta_n \rightarrow 0$ in probability, as $n \rightarrow \infty$, we may conclude that

$$\mathbb{P} \{ \mathbb{T}_n \leq t \} \approx \mathbb{P}^* \{ \mathbb{T}_n^* \leq t \}$$

in probability, for n large enough. This justifies the bootstrap quantile approximation.

REMARK. This test is consistent. Indeed, under the null, observe that

$$\mathbb{T}_n \leq L_n \sup_B |\mathbb{Z}_n(B)| = O_p(L_n).$$

Although this diverges, we can always estimate the distribution of \mathbb{T}_n for n large enough. Under the alternative hypothesis, $H_A : C = C_1$ for a fixed $C_1 \neq C_0$, we have that

$$\mathbb{T}_n \geq \sqrt{n} |C_0(B) - C_1(B)| - |\mathbb{Z}_n(B)| = O_p(n^{1/2})$$

for any rectangle B where C_0 and C_1 differ. Such a rectangle exists under the alternative and the increasing sequence \mathcal{F}_n likely contains at least one such rectangle for relatively small n . The improved power of our test statistic is confirmed in our simulation study.

REMARK. Taking the supremum over all families of size L_n makes the computation of the statistic \mathbb{T}_n non-trivial. If we do not take the supremum over all disjoint rectangles B_1, \dots, B_{L_n} in (4), but rather limit ourselves to a particular sequence of families Π_n with $B_{n1}, \dots, B_{nL_n} \in \Pi_n$, then L_n is allowed to grow at a faster power rate n^ζ , $\zeta < 1/4$ in the sample size n . Again, although the resulting statistic $\mathbb{S}_n = \sum_{k=1}^{L_n} |\mathbb{Z}_n(B_{nk})|$ does not have a weak limit for $L_n \rightarrow \infty$, its distribution can be

consistently estimated by a bootstrap estimator. This test is less powerful than the ATV test, but \mathbb{S}_n is easier to compute than \mathbb{T}_n . We will report on this test elsewhere.

3. PARAMETRIC HYPOTHESIS

In this section we consider the problem of testing if the underlying copula C belongs to a parametric family $\mathcal{C} := \{C_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$. That is, the null hypothesis states that $C = C_{\theta_0}$ for some $\theta_0 \in \Theta$. Suppose that we have a consistent estimator $\hat{\theta} = \hat{\theta}(\mathbb{H}_n)$ of θ_0 .

Replacing C_0 by $C_{\hat{\theta}}$ in the definition of the test statistic \mathbb{T}_n , we consider the process

$$(24) \quad \mathbb{Y}_n = \sqrt{n}(\mathbb{C}_n - C_{\hat{\theta}}) = \mathbb{Z}_n - \sqrt{n}(C_{\hat{\theta}} - C),$$

and its bootstrap version

$$(25) \quad \mathbb{Y}_n^* = \mathbb{Z}_n^* - \sqrt{n}(C_{\hat{\theta}^*} - C_{\hat{\theta}}).$$

based on the *non-parametric* bootstrap estimate $\hat{\theta}^* = \hat{\theta}(\mathbb{H}_n^*)$, obtained after resampling with replacement from the original sample. Note that

$$(26) \quad \mathbb{Y}_n^* = \sqrt{n}(\mathbb{C}_n^* - C_{\hat{\theta}^*}) - \sqrt{n}(\mathbb{C}_n - C_{\hat{\theta}}) \neq \sqrt{n}(\mathbb{C}_n^* - C_{\hat{\theta}}).$$

Indeed, the process $\sqrt{n}(\mathbb{C}_n^* - C_{\hat{\theta}^*})$, while perhaps a natural candidate, does not yield a consistent estimate of the distribution of \mathbb{Z}_n .

We stress that our approach does not involve the *parametric* bootstrap, as studied by Genest and Rémillard (2008) to estimate the limiting law of copula-based statistics. In other words, we calculate $\hat{\theta}^*$ after resampling from the empirical distribution \mathbb{H}_n , and not from the law given by the parametric copula $C_{\hat{\theta}}$.

We impose some regularity on our parameter estimate $\hat{\theta}$.

(C3) There exists a $\psi : \mathbb{R}^2 \mapsto \mathbb{R}^p$ with $\int \|\psi\|^4 dH < \infty$ such that

$$\hat{\theta} - \theta_0 = \int \psi d(\mathbb{H}_n - H) + \varepsilon_n,$$

and

$$\widehat{\theta}^* - \widehat{\theta} = \int \psi d(\mathbb{H}_n^* - \mathbb{H}_n) + \varepsilon_n^*,$$

under the null hypothesis, with $\|\varepsilon_n\| = o_p(n^{-1/2}/L_n)$ and $\|\varepsilon_n^*\| = o_{p^*}(n^{-1/2}/L_n)$ in probability.

Note that the estimators satisfying (C3) are closely related to the estimators in the class \mathcal{R} of regular estimators, as defined by Genest and Rémillard (2008).

EXAMPLE (Estimators based on the inversion of Kendall's tau). As an example, we verify condition (C3) for estimators based on the inversion of Kendall's tau. Let $\theta = g(\tau)$ for some twice differentiable function g and Kendall's $\tau := 4\mathbb{E}[C_\theta(U, V)] - 1$, with the expectation taken over $(U, V) \sim C_\theta$. Kendall's τ is estimated empirically by

$$\widehat{\tau}_n := \frac{4}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n 1 \{(Y_j - Y_i)(X_j - X_i) > 0\} - 1.$$

Then $U_n := \widehat{\tau}_n + 1$ is a U-statistics of order 2 for the kernel

$$h((x_1, y_1); (x_2, y_2)) = 2 \cdot 1\{(y_2 - y_1)(x_2 - x_1) > 0\}.$$

The projection of $U_n - \mathbb{E}[U_n]$ onto the space of all statistics of the form $\sum_{i=1}^n g_i(X_i, Y_i)$, for arbitrary measurable functions g_i with $\mathbb{E}[g_i(X, Y)] < \infty$, is

$$\widehat{U}_n = \sum_{i=1}^n \mathbb{E}[U_n - \mathbb{E}[U_n] | X_i] = \frac{2}{n} \sum_{i=1}^n h_1(X_i, Y_i)$$

with

$$h_1(x, y) = \{F^l(x, y) + F^r(x, y) - \mathbb{E}[F^l(X, Y) + F^r(X, Y)]\}$$

and $F^l(x, y) = P(X < x, Y < y)$ and $F^r(x, y) = P(X > x, Y > y)$. By Hajék's projection principle,

$$\text{Var}(U_n - \widehat{U}_n) = \text{Var}(U_n) - \text{Var}(\widehat{U}_n).$$

From the proof of Theorem 12.3 in Van der Vaart (1998), due to Hoeffding (1948), the difference is of order $O(1/n^2)$. Consequently, $U_n - \mathbb{E}[U_n] = \widehat{U}_n + R_n$ with $R_n =$

$O_p(1/n)$ so that

$$\widehat{\tau}_n - \tau = U_n - \mathbb{E}[U_n] = \widehat{U}_n + R_n = \frac{2}{n} \sum_{i=1}^n h_1(X_i, Y_i) + O_p(1/n).$$

Hence, if g is twice continuously differentiable in the neighborhood of τ , a limited expansion ensures that $\widehat{\theta} := g(\widehat{\tau}_n)$ satisfies the first part of (C3). The second (bootstrap) part of (C3) follows from the same reasoning.

Moreover, we need more regularity concerning $\theta \mapsto C_\theta$ itself.

(C4) For every $(s, t) \in [0, 1]^2$, the function $\theta \mapsto C_\theta(s, t)$ has continuous partial derivatives $\dot{C}_\theta(s, t) = (\partial/\partial\theta)C_\theta(s, t)$ that satisfy a Hölder condition with Hölder exponent $\nu > 0$ locally: there exists a constant $K < \infty$ such that

$$\sup_{s, t} \|\dot{C}_\theta(s, t) - \dot{C}_{\theta_0}(s, t)\| \leq K \|\theta - \theta_0\|^\nu,$$

for every θ in a neighborhood of θ_0 . Moreover, \dot{C}_{θ_0} is of bounded variation.

The regularity condition (C4) is satisfied for most of standard copula families. Simple calculations show that it is the case for the Gaussian-, Clayton- and the Frank-copula families in particular. Although copula derivatives $\partial C_\theta(x, y)/\partial x$ and $\partial C_\theta(x, y)/\partial y$ with respect to their arguments often exhibit discontinuities or non-existence near their boundaries³, the derivatives $\partial C_\theta(x, y)/\partial\theta$ with respect to the copula parameter θ behave a lot more regularly.

THEOREM 3. *Let $\mathbb{Y}_n = \{\mathbb{Y}_n(f), f \in \mathcal{F}_n\}$ and $\mathbb{Y}_n^* = \{\mathbb{Y}_n^*(f), f \in \mathcal{F}_n\}$ with \mathcal{F}_n in (10) as defined above. Assume that conditions (C1), (C2), (C3) and (C4) hold. Then, under the null hypothesis $H_0 : C = C_\theta, \theta \in \Theta$, we have*

$$(27) \quad \lim_{n \rightarrow \infty} \mathbb{E}[d_{BL_1}(\mathbb{Y}_n, \mathbb{Y}_n^*)] = 0.$$

This result implies that the distribution of the test statistic

$$(28) \quad \widehat{\mathbb{T}}_n = \sup_{f \in \mathcal{F}_n} |\mathbb{Y}_n(f)| = \sup_{B_1, \dots, B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Y}_n(B_k)|$$

³ justifying conditions such as (C1), see Segers (2012).

can be “bootstrapped” by the distribution of

$$(29) \quad \widehat{\mathbb{T}}_n^* = \sup_{f \in \mathcal{F}_n} |\mathbb{Y}_n^*(f)| = \sup_{B_1, \dots, B_{L_n}} \sum_{k=1}^{L_n} |\mathbb{Y}_n^*(B_k)|.$$

COROLLARY 4. *Assume that conditions (C1), (C2), (C3) and (C4) hold. Then, under the null hypothesis $H_0 : C = C_\theta$, $\theta \in \Theta$,*

$$(30) \quad \lim_{n \rightarrow \infty} \mathbb{E} \left[d_{BL_1}(\widehat{\mathbb{T}}_n, \widehat{\mathbb{T}}_n^*) \right] = 0.$$

Often, (C3) can be replaced by

(C3') There exists a $\psi : \mathbb{R}^2 \mapsto \mathbb{R}^p$ with $\int \|\psi\|^4 dC < \infty$ such that

$$\widehat{\theta} - \theta_0 = \frac{1}{n} \sum_{i=1}^n \{ \psi(\mathbb{F}_n(X_i), \mathbb{G}_n(Y_i)) - \mathbb{E}[\psi(F(X_i), G(Y_i))] \} + \varepsilon_n,$$

and

$$\widehat{\theta}^* - \widehat{\theta} = \frac{1}{n} \sum_{i=1}^n \{ \psi(\mathbb{F}_n^*(X_i^*), \mathbb{G}_n^*(Y_i^*)) - \psi(\mathbb{F}_n(X_i), \mathbb{G}_n(Y_i)) \} + \varepsilon_n^*,$$

under the null hypothesis, with $\|\varepsilon_n\| = o_p(n^{-1/2}/L_n)$ and $\|\varepsilon_n^*\| = o_{p^*}(n^{-1/2}/L_n)$ in probability.

This is a consequence of the following result.

PROPOSITION 5. *Assume (C1) holds. Any estimator $\widehat{\theta}$ satisfying (C3'), satisfies (C3).*

Copula parameters are typically estimated through pseudo-observations or ranks, without any assumption on the marginal distributions. For this reason the copula estimators that satisfy (C3') are relevant. They are very closely related to the estimators in the class \mathcal{R}_1 of Genest and Rémillard (2008). In particular, the maximum pseudo-likelihood estimator, that maximizes the pseudo log-likelihood function $\int \log c_\theta d\mathbb{C}_n$ over $\theta \in \Theta$, see, for instance, Genest et al. (1995) or Shih and Louis (1995), satisfies (C3') under suitable regularity conditions on the copula density $c_\theta(u, v)$.

Since the bootstrapped copula process \mathbb{Y}_n^* is new, it is noteworthy to stress that it provides a valuable alternative to the usual parametric bootstrap. Now, assume $L_n = L$ is a constant, to retrieve the standard framework.

COROLLARY 6. *Assume that conditions (C1), (C3) and (C4) hold. Then, the process $\{\mathbb{Y}_n(s, t), 0 \leq s, t \leq 1\}$ tends weakly towards a Gaussian process in $\ell^\infty([0, 1]^2)$. Moreover, the bootstrapped process $\{\mathbb{Y}_n^*(s, t), 0 \leq s, t \leq 1\}$ converges weakly to the same Gaussian process in probability in $\ell^\infty([0, 1]^2)$.*

4. APPLICATIONS AND NUMERICAL STUDIES

We present a limited numerical study which should serve as a proof of principle rather than the final word on this subject. The evaluation of GOF tests in copula settings is a complex problem and only partial answers can be found in literature: see the surveys of Berg (2009), Genest et al. (2009) and, more recently, Fermanian (2012). A full-scale numerical analysis is beyond the scope of this paper.

4.1. Heuristics. For two copula densities c_0 and c_1 , we define the *difference* sets A^+ and A^- as

$$A^+ = \{(s, t) : c_0(s, t) > c_1(s, t)\}, \text{ and } A^- = \{(s, t) : c_0(s, t) < c_1(s, t)\}.$$

We assume that there exist minimal integers M^+ and M^- such that

$$A^+ = \bigcup_{i=1}^{M^+} D_i^+ + A_0^+ \text{ and } A^- = \bigcup_{i=1}^{M^-} D_i^- + A_0^-,$$

for Lebesgue null sets A_0^+ and A_0^- and open sets D_i^+ and D_i^- in $[0, 1]^2$. The proposed statistic \mathbb{T}_n is designed to sample L_n rectangles in order to maximize the difference between the “true” and postulated copulas. In situations where the geometry of the difference sets A^+ and A^- is complex (most often when M^+ and M^- are large), \mathbb{T}_n can “pick out” the disjoint regions D_i^+ and D_i^- , and one could expect superior performances consequently. Moreover, since we typically do not know the values of M^+ and M^- , it makes sense to let $L_n \rightarrow \infty$, which is exactly what our theoretical results allow us to do.

However, in situations where M^+ and M^- are small or one of the regions D_i is much larger than the others, it is not advantageous to sample numerous disjoint rectangles.

In these situations, just a single well placed rectangle can pick essentially all the mass of sets A^+ or A^- , while the remaining $L_n - 1$ rectangles are just collecting noise and consequently diminish the power of the statistic \mathbb{T}_n .

Most common scenarios encountered in the literature compare Frank, Clayton, Gumbel, and Gauss copulas with each other, after controlling for some dependence indicator (typically Kendall's tau): see, for instance, Berg (2009), Genest and Rémilliard (2008) and Genest et al. (2009). However, all these pairings produce trivial difference sets A^+ and A^- , as revealed in the contour plots and 3D plots of $c_0 - c_1$ of Figure 1. We see that nearly all the mass difference between copula densities c_0 and c_1 is concentrated in a single spot, located in either the lower left or upper right corner⁴. Therefore, these common simulation scenarios are tailored towards many standard GOF tests such as KS and CvM tests.

We are not aware of any argument that justifies such specific types of pairing, except for analytical tractability. Figures 2 and 3, however, paint a very different scenario with more elaborate difference sets A^+ and A^- that appears in real life situations. How often and to what extent this complex situation is encountered is still an open issue.

In this study, the copula densities c_1 were estimated by kernel density estimators based on the following data:

- The bivariate *ARCH* process $(X_1, Y_1), \dots, (X_n, Y_n)$, with $n = 10^6$, was generated as follows: First, we created independent $Z_i \sim N(0, 1)$ and $W_i = Z_i(1 + 0.6W_{i-1}^2)^{1/2}$, with $W_0 = 0$. Second, we set $(X_i, Y_i) := (W_{100i}, W_{100i+1})$, creating nearly independent observations.
- The *Mixture Copula* data $(X_1, Y_1), \dots, (X_n, Y_n)$, with $n = 10^6$, are generated from the mixture $c_0(s, t) = \frac{1}{2}c_F(s, t) + \frac{1}{2}c_F(1 - s, t)$ for the Frank copula c_F with Kendall's $\tau = 0.4$.
- The *Euro-Dollar* data $(X_1, Y_1), \dots, (X_n, Y_n)$, with $n = 1800$, are quoted currency exchange values. X is the daily percentage change of the Euro against

⁴Here Kendall's $\tau = 0.4$, but we observed similar plots for different values of τ .

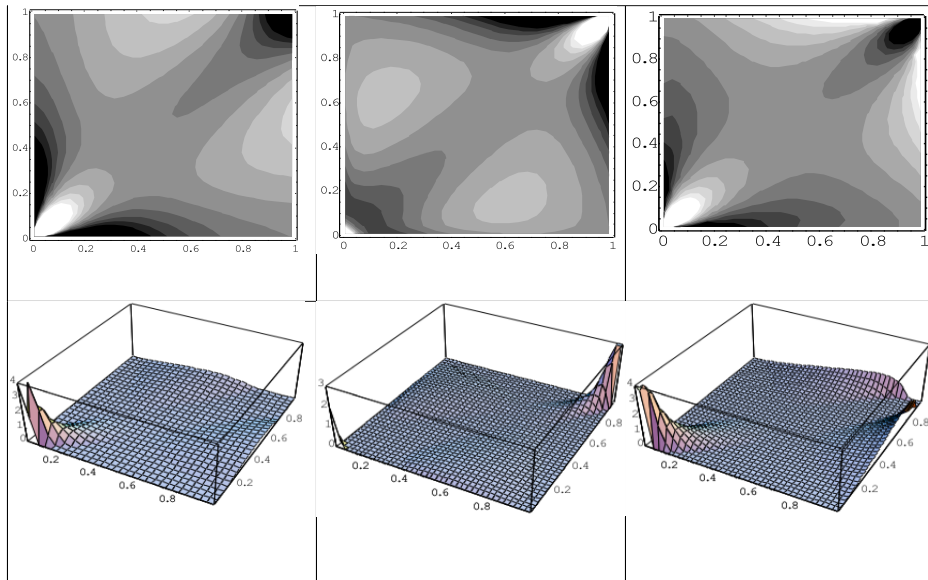


FIGURE 1. Contour and 3D plots of common comparisons: Clayton-Frank (left panel); Gumbel-Frank (middle panel); Clayton-Gumbel (right panel).

the US dollar, while Y corresponds to the daily change of the Canadian dollar against the US dollar.

- The *Silver-Gold* data $(X_1, Y_1), \dots, (X_n, Y_n)$, with $n = 5000$, presents the log ratio of the average daily price of silver and gold futures respectively. For instance, $X_i = \log(S_{i+1}/S_i)$ based on the average price S_i of silver in US dollars on day i .

In the case of *Mixture copula* and *ARCH*, we used $c_0(s, t) = 1$ (the independence copula). In the case of *Euro-Dollar* and *Silver-Gold*, we took the Frank copula density with parameters $\tau = 2.6$ and $\tau = 3.4$, respectively, for c_0 . The latter parameters were chosen after minimizing the (estimated) L_1 -distance between c_0 and c_1 . The difference sets are easily depicted by dark and bright sections of the contour plots and the 3D plots clearly indicate that the mass difference between copula densities

c_0 and c_1 is not concentrated in a single spot.

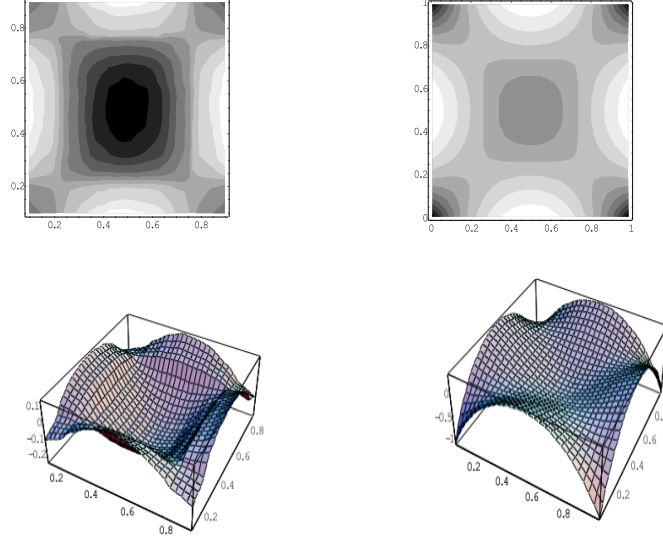


FIGURE 2. Contour and 3D plots of complex comparisons based on synthetic data: ARCH (left panel); Mixture Copulas (right panel).

4.2. GOF tests in practice. We generated the data sets *ARCH* and *Mixture Copula* as described above. For each data set, we run two sets of simulations:

- (ARCH-S and Mixture-S) Test the simple null hypothesis $C_0(s, t) = st$ using the methodology of Section 2.
- (ARCH-C and Mixture-C) Test the composite null hypothesis that C_0 is a Frank copula using the procedure described in Section 3.

In both cases, the null hypothesis is wrong and should be rejected.

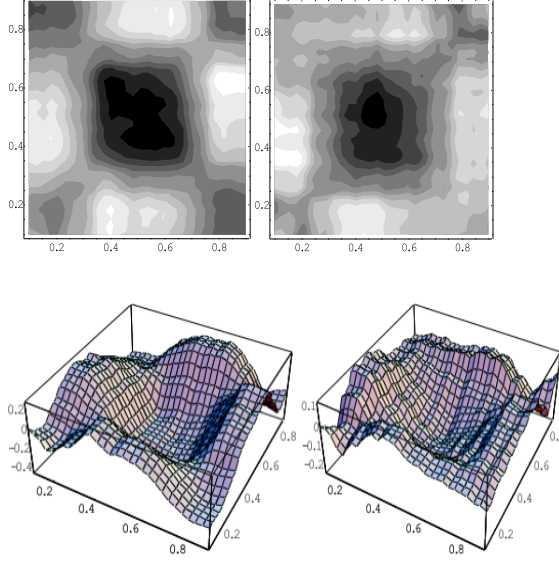


FIGURE 3. Contour and 3D plots of complex comparisons based on real data: Euro-Dollar (left panel); Silver-Gold (right panel).

In our simulations, the statistics \mathbb{T}_n and $\hat{\mathbb{T}}_n$ are computed using the number of rectangles $L_n = \lfloor \ln(n) \rfloor - 2$. In our experience $\lfloor \ln(n) \rfloor - 3$ or $\lfloor \ln(n) \rfloor - 1$ worked equally well. However, we did encounter some performance issues with larger choices of L_n , for instance $L_n = n^{1/3}$ or $L_n = n^{1/2}$. We approximated the p -values of our statistics via the bootstrap procedures introduced in sections 2 and 3. For each approximation, we used 1,000 bootstrap samples. For the second set of simulations (ARCH-C and Mixture-C), we computed the parameters $\hat{\theta}$ and $\hat{\theta}^*$ by the usual pseudo-maximum likelihood procedure. Each procedure is repeated 100 times. We report the percentage of times that the computed p -value is below $\alpha = 0.05$.

Although the computation of ATV statistic consists merely of a complex optimization problem, we found that the following brute force approach yields good results:

- (1) sample random disjoint rectangle B_1, \dots, B_{L_n} ;
- (2) compute the quantity $A = \sum_{i=1}^{L_n} |\mathbb{Z}_n(B_i)|$ or $\sum_{i=1}^{L_n} |\mathbb{Y}_n(B_i)|$;
- (3) repeat this procedure $3000 \cdot L_n$ times and record the largest quantity A .

We implemented two simple computational tricks to assist the above scheme in order to speed up the procedure.

- Since randomly chosen rectangles are often overlapping, set $a := 1/2$ and choose rectangle B_1 at random such that $\text{area}(B_1) < a$. Then, set $a := a - \text{area}(B_1)/2$. Given B_1, \dots, B_k rectangles, sample a rectangle B_{k+1} such that $\text{area}(B_{k+1}) < a$ and $B_{k+1} \cap B_j = \emptyset$ for $j = 1, \dots, k$. Set $a := a - \text{area}(B_{k+1})/2$ and repeat the procedure until $k = L_n$.
- Pre-compute and store the matrix $\mathbb{C}_n(i/n, j/n)$, $1 \leq i, j \leq n$. We can simply recall the appropriate $\mathbb{C}_n(i/n, j/n)$ each time we need to compute $\mathbb{C}_n(s, t)$.⁵

The above computations with 1000 bootstrap samples took 10 – 15 seconds on a standard PC. More sophisticated optimization methods, like the Accelerated Random Search (Appel et al., 2004), yield much faster computation (less than 1 second). Our limited numerical study confirms the above assessment. Table 1 shows that the ATV test outperforms the KS and CvM tests in the case of complex pairing, while Table 2 confirms that the ATV test is inferior in case of the commonly used pairings of Figure 1. In Table 2, for each pair of copulas, say Clayton - Frank, we generated n observations from the first copula (Clayton), and we tested the null hypothesis that the second copula (Frank) is the true underlying copula. Table 3 shows that the significance level of the ATV test is below 0.05. The data was simulated from the null hypothesis. In all tables, Kendall's $\tau = 0.4$.

⁵Indeed, naive computation of \mathbb{C}_n requires n operations. For our optimization scheme, this would lead to $3000 \cdot n \cdot L_n$ operations. In comparison, using recursion, the grid can be computed with n^2 operations and only once, resulting with the total of $n^2 + 3000L_n$ operations.

type	n	ARCH-S	ARCH-C	Rotated-S	Rotated -C
ATV	400	75%	80%	41%	25%
KS	400	6%	4%	8%	12%
CvM	400	25%	50%	6%	15%
ATV	800	100%	99%	94%	98%
KS	800	32%	50%	20%	25%
CvM	800	50%	92%	31%	84%

TABLE 1. (Complex pairing related to Figure 2) Relative frequencies of rejected null hypotheses under $\alpha = 0.05$.

type	n	Clayton - Frank	Gumbel - Frank	Clayton - Gumbel
ATV	400	42%	26%	88%
KS	400	58%	25%	90%
CvM	400	84%	47%	95%
ATV	800	92%	58%	94%
KS	800	98%	53%	98%
CvM	800	100%	73%	100%

TABLE 2. (Trivial pairing related to Figure 1) Relative frequencies of rejected null hypotheses under $\alpha = 0.05$.

type	n	Clayton - Clayton	Gumbel - Gumbel	Frank -Frank
ATV	400	3%	2%	2%
KS	400	4%	5%	4%
CvM	400	4%	5%	4%
ATV	800	2%	4%	3%
KS	800	3%	3%	5%
CvM	800	5%	3%	6%

TABLE 3. (Errors of the first kind) Relative frequencies of rejected null hypotheses under $\alpha = 0.05$.

5. PROOFS

Throughout the proofs, we assume without loss of generality that $F = G = I$ (uniform marginal distributions). This implies that $H = C$. This is justified by the following lemma.

LEMMA 7. *Let F, G be continuous distribution functions. Denote by \tilde{H} the cdf of $(F(X), G(Y))$ and by \tilde{C} its associated copula. The empirical copula associated to the sample $(F(X_1), G(Y_1)), \dots, (F(X_n), G(Y_n))$ is denoted by $\tilde{\mathbb{C}}_n$. We have*

$$C(x, y) = \tilde{C}(x, y) = \tilde{H}(x, y) \text{ for all } x, y \in [0, 1].$$

Moreover,

$$\mathbb{C}_n\left(\frac{i}{n}, \frac{j}{n}\right) = \tilde{\mathbb{C}}_n\left(\frac{i}{n}, \frac{j}{n}\right) \text{ for } i, j = 0, 1, \dots, n.$$

Proof. See Lemma 1 in Fermanian et al. (2004). □

Since the letter C is reserved for the copula function, we use the letters K, K_0, K_1 , etc. in the sequel to denote generic constants.

5.1. PROOF OF PRELIMINARY RESULTS. In general, note that, for each $f \in \mathcal{F}_n$ defined in (10), we can write

$$\mathbb{Z}_n(f) = \sum_{k=1}^{L_n} c_k \mathbb{Z}_n(B_k) = \sum_{j=1}^{4L_n} \sigma_j \mathbb{Z}_n(s_j, t_j)$$

and

$$\mathbb{Z}_n^*(f) = \sum_{j=1}^{4L_n} \sigma_k \mathbb{Z}_n^*(s_j, t_j)$$

for some $\sigma_j \in \{-1, +1\}$ and $0 \leq s_j, t_j \leq 1$ using formula (5). Let

$$\alpha_n \equiv \sqrt{n}(\mathbb{H}_n - H)$$

be the ordinary empirical process in $[0, 1]^2$, and let its oscillation modulus be defined as

(31)

$$\mathbb{M}_n(\delta) := \sup \{ |\alpha_n(s, t) - \alpha_n(s', t')| : |s - s'| \leq \delta, |t - t'| \leq \delta, 0 \leq s, s', t, t' \leq 1 \},$$

for any $\delta > 0$.

LEMMA 8. *Let $(\delta_n)_{n \geq 0}$ be a sequence of positive real numbers such that $n\delta_n/\log n \rightarrow \infty$. Then we have*

$$\mathbb{M}_n(\delta_n) = O(\delta_n^{1/2}(\log n)^{1/2}) \quad \text{almost surely.}$$

Proof. We apply Proposition 13 with $\lambda_n = K_0 \delta_n^{1/2} (\log n)^{1/2}$ for some constant $K_0 > 0$. Since $n^{-1/2} \lambda_n / \delta_n = K_0 (\log n / (n \delta_n))^{1/2}$ tends to zero, this inequality can be rewritten

$$\mathbb{P} \{ \mathbb{M}_n(\delta_n) > \lambda_n \} \leq \frac{K_1}{\delta_n} \exp \left(-\frac{K_2 \psi(1) \lambda_n^2}{\delta_n} \right) = K_1 n \exp \left(-K_2 K_0^2 \psi(1) \log n \right),$$

for some constants K_1, K_2 and n sufficiently large. When K_0 is sufficiently large, we check that

$$\mathbb{P} \{ \mathbb{M}_n(\delta_n) > \lambda_n \} \leq \frac{K_3}{n^2},$$

for some constant K_3 . Invoke the Borel-Cantelli Lemma to conclude the proof. \square

In addition, set

$$(32) \quad \tilde{\mathbb{Z}}_n(s, t) = \alpha_n(s, t) - C_1(s, t) \alpha_n(s, 1) - C_2(s, t) \alpha_n(1, t),$$

and define

$$\tilde{\mathbb{Z}}_n(f) = \sum_{k=1}^{4L_n} \sigma_k \tilde{\mathbb{Z}}_n(s_k, t_k).$$

PROPOSITION 9. *Under conditions (C1) and (C2), we have*

$$\lim_{n \rightarrow \infty} d_{BL_1}(\mathbb{Z}_n, \tilde{\mathbb{Z}}_n) = 0.$$

Proof. First, we observe that

$$\begin{aligned} d_{BL_1}(\mathbb{Z}_n, \tilde{\mathbb{Z}}_n) &= \sup_{h \in BL_1(\ell^\infty(\mathcal{F}_n))} \left| \mathbb{E}[h(\mathbb{Z}_n) - h(\tilde{\mathbb{Z}}_n)] \right| \\ &\leq \delta + 2\mathbb{P} \left\{ \sup_{f \in \mathcal{F}_n} |\mathbb{Z}_n(f) - \tilde{\mathbb{Z}}_n(f)| > \delta \right\}. \end{aligned}$$

The latter inequality holds for any $\delta > 0$, and uses the fact that $|h|$ is bounded by 1 and has Lipschitz constant 1. It remains to show that

$$\sup_{f \in \mathcal{F}_n} |\mathbb{Z}_n(f) - \tilde{\mathbb{Z}}_n(f)| \rightarrow 0,$$

in probability, as $n \rightarrow \infty$. Second, we note that

$$\sup_{f \in \mathcal{F}_n} |\mathbb{Z}_n(f) - \tilde{\mathbb{Z}}_n(f)| \leq 4L_n \sup_{0 \leq s, t \leq 1} |\mathbb{Z}_n(s, t) - \tilde{\mathbb{Z}}_n(s, t)|.$$

The remainder of the proof generalizes Proposition 4.2 of Segers (2012). Now,

$$\sup_{0 \leq s, t \leq 1} |\mathbb{Z}_n(s, t) - \tilde{\mathbb{Z}}_n(s, t)| \leq I + II$$

with

$$I = \sup_{0 \leq s, t \leq 1} |\alpha_n(\mathbb{F}_n^- s, \mathbb{G}_n^- t) - \alpha_n(s, t)|$$

and

$$II = \sup_{0 \leq s, t \leq 1} \left| \sqrt{n} [C(\mathbb{F}_n^- s, \mathbb{G}_n^- t) - C(s, t)] + C_1(s, t)\alpha_n(s, 1) + C_2(s, t)\alpha_n(1, t) \right|.$$

The first term, I , can be bounded as follows. Set $\beta_{n1}(s) = \sqrt{n}(\mathbb{F}_n^- s - s)$ and $\beta_{n2}(t) = \sqrt{n}(\mathbb{G}_n^- t - t)$. By the Chung-Smirnov LIL, we have

$$\sup_{0 \leq s, t \leq 1} |\beta_{n1}(s)| + |\beta_{n2}(t)| = O((\log \log n)^{1/2}) \quad \text{almost surely.}$$

Using Lemma 8 with $\delta = n^{-1/2}(\log \log n)^{1/2}$, we get

$$\sup_{|s-s'|+|t-t'| < \delta} |\alpha_n(s, t) - \alpha_n(s', t')| = O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}),$$

almost surely. This implies that $I = O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4})$, almost surely.

For the second term, we note that

$$\begin{aligned} II &= \sup_{0 \leq s, t \leq 1} \left| \sqrt{n} [C(\mathbb{F}_n^- s, \mathbb{G}_n^- t) - C(s, t)] + C_1(s, t)\alpha_n(s, 1) + C_2(s, t)\alpha_n(1, t) \right| \\ &\leq IIa + IIb + IIc, \end{aligned}$$

with

$$\begin{aligned} IIa &= \sup_{s, t \notin (\varepsilon_n, 1-\varepsilon_n)} \left| \sqrt{n} [C(\mathbb{F}_n^- s, \mathbb{G}_n^- t) - C(s, t)] + C_1(s, t)\alpha_n(s, 1) + C_2(s, t)\alpha_n(1, t) \right|, \\ IIb &= \sup_{\varepsilon_n < s, t < 1-\varepsilon_n} |C_1(s_n, t_n)\beta_{n1}(s) + C_1(s, t)\alpha_n(s, 1)|, \\ IIc &= \sup_{\varepsilon_n < s, t < 1-\varepsilon_n} |C_2(s'_n, t'_n)\beta_{n2}(t) + C_2(s, t)\alpha_n(1, t)| \end{aligned}$$

Here $\varepsilon_n \rightarrow 0$ will be specified later, and $|s_n - s| \leq n^{-1/2}|\beta_{n1}(s)|$, $|t_n - t| \leq n^{-1/2}|\beta_{n2}(t)|$, $|s'_n - s| \leq n^{-1/2}|\beta_{n1}(s)|$ and $|t'_n - t| \leq n^{-1/2}|\beta_{n2}(t)|$.

Let us bound IIa . Since $|C_1| \leq 1$ and $|C_2| \leq 1$ as copulas are Lipschitz with Lipschitz constant 1, we get by the mean value theorem that

$$IIa \leq \sup_{s,t \in [0, \varepsilon_n] \cup [1-\varepsilon_n, 1]} [|\beta_{n1}(s)| + |\beta_{n2}(t)| + |\alpha_n(s, 1)| + |\alpha_n(1, t)|].$$

The Bahadur-Kiefer theorem (Shorack and Wellner 2009, p. 585) states that

$$\sup_{0 \leq s \leq 1} |\beta_{n1}(s) + \alpha_n(s, 1)| = O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}) \quad \text{almost surely.}$$

Furthermore,

$$\begin{aligned} \varepsilon_n^{-1/2} \sup_{s \notin (\varepsilon_n, 1-\varepsilon_n)} |\alpha_n(s, 1)| &\leq \sup_{s \notin (\varepsilon_n, 1-\varepsilon_n)} s^{-1/2}(1-s)^{-1/2} |\alpha_n(s, 1)| \\ &= (\log n)^{1/2} \quad \text{almost surely,} \end{aligned}$$

see Theorem 2 in Mason (1981). (Alternatively, we can apply Lemma 8 for $n\varepsilon_n/\log n \rightarrow \infty$.) Combining all these bounds entails then

$$IIa = O(\varepsilon_n^{1/2}(\log n)^{1/2} + n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}) \quad \text{almost surely.}$$

Next, we bound IIb as follows.

$$\begin{aligned} IIb &= \sup_{\varepsilon_n < s, t < 1-\varepsilon_n} |C_1(s_n, t_n)\beta_{n1}(s) + C_1(s, t)\alpha_n(s, 1)| \\ &\leq \sup_{\varepsilon_n < s, t < 1-\varepsilon_n} |\beta_{n1}(s) + \alpha_n(s, 1)| + \sup_{\varepsilon_n < s, t < 1-\varepsilon_n} |C_1(s_n, t_n) - C_1(s, t)| |\alpha_n(s, 1)|. \end{aligned}$$

The first term on the right is $O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4})$ almost surely, as we have seen before. For the second term, using condition (C1) we have, for some $r > 0$, $\beta \geq 0$ and $K < \infty$,

$$\begin{aligned} &\sup_{\varepsilon_n < s, t < 1-\varepsilon_n} |C_1(s_n, t_n) - C_1(s, t)| |\alpha_n(s, 1)| \leq \\ &K \sup_{\varepsilon_n < s, t < 1-\varepsilon_n} (|s_n - s|^r + |t_n - t|^r) (s^{-\beta}(1-s)^{-\beta} + s_n^{-\beta}(1-s_n)^{-\beta}) |\alpha_n(s, 1)|. \end{aligned}$$

Since

$$s_n = s \left(1 + \frac{s_n - s}{s} \right) \geq s \left(1 - \frac{|s_n - s|}{\varepsilon_n} \right) \geq \frac{s}{2},$$

and

$$1 - s_n \geq (1 - s) \left(1 - \frac{|s_n - s|}{\varepsilon_n} \right) \geq \frac{1 - s}{2},$$

almost surely and for n sufficiently large, for all $\varepsilon_n \rightarrow 0$ and $n\varepsilon_n^2/\log n \rightarrow \infty$. We deduce, using Theorem 2 in Mason (1981),

$$\begin{aligned} & \sup_{\varepsilon_n < s, t < 1 - \varepsilon_n} |C_1(s_n, t_n) - C_1(s, t)| |\alpha_n(s, 1)| \\ & \leq \sup_{\varepsilon_n < s, t < 1 - \varepsilon_n} K(|s_n - s|^r + |t_n - t|^r) s^{\frac{1}{2} - \beta} (1 - s)^{\frac{1}{2} - \beta} (\log n)^{1/2} \\ & \leq K n^{-r/2} \max(\varepsilon_n^{\frac{1}{2} - \beta}, 1) (\log n)^{1/2} \sup_{\varepsilon_n < s, t < 1 - \varepsilon_n} (|\beta_{n1}(s)|^r + |\beta_{n2}(t)|^r) \\ & = O \left(n^{-r/2} (\log n)^{1/2} (\log \log n)^{r/2} \max(\varepsilon_n^{\frac{1-2\beta}{2}}, 1) \right) \end{aligned}$$

almost surely. We now specify the choice of $\varepsilon_n = n^{-p}$ further, with p depending on β and r . If $2\beta > 2r + 1$, we take $0 < p < r/(2\beta - 1)$. Otherwise, we take $0 < p < 1/2$. For both cases, these choices ensure that $IIb = O(n^{-q})$ almost surely, for some $q > 0$. The same bound applies to IIc . Since $L_n = o(\log n)$ by assumption (C2), we obtain $L_n(I + II) \rightarrow 0$, almost surely, as $n \rightarrow \infty$, and the proof is complete. \square

Next, we turn our attention to the bootstrap counterparts. We define $\alpha_n^*(s, t) = \sqrt{n}(\mathbb{H}_n^* - \mathbb{H}_n)(s, t)$ as the ordinary bootstrap empirical process in $[0, 1]^2$. We prove the following exponential inequality for the oscillation

$$\mathbb{M}_n^*(\delta) = \sup_{|s-s'| < \delta, |t-t'| < \delta} |\alpha_n^*(s, t) - \alpha_n^*(s', t')|$$

of this process.

LEMMA 10. *For all bounded sequences δ_n such that $n\delta_n/\log(n) \rightarrow \infty$ as $n \rightarrow \infty$,*

$$(33) \quad \mathbb{M}_n^*(\delta_n) = O(\delta_n^{1/2} (\log n)^{1/2}) \quad \text{almost surely.}$$

Note that the sequence (δ_n) may be constant.

Proof. Since α_n^* is a step function, we find that

$$\sup_{|s-s'| < \delta_n, |t-t'| < \delta_n} |\alpha_n^*(s, t) - \alpha_n^*(s', t')|$$

equals

$$\max_{i,i',j,j' \in \{1,\dots,n\}, |X_i - X_{i'}| < \delta_n, |Y_j - Y_{j'}| < \delta_n} |\alpha_n^*(X_i, Y_j) - \alpha_n^*(X_{i'}, Y_{j'})|.$$

Let us rewrite

$$|\alpha_n^*(X_i, Y_j) - \alpha_n^*(X_{i'}, Y_{j'})| = n^{-1/2} \sum_{k=1}^n \{V_{kijij'} - \mathbb{E}^*[V_{kijij'}]\},$$

as a sum of bounded independent random variables with

$$V_{kijij'} := 1\{X_k^* \leq X_i, Y_k^* \leq Y_j\} - 1\{X_k^* \leq X_{i'}, Y_k^* \leq Y_{j'}\},$$

conditionally on the sample $(X_1, Y_1), \dots, (X_n, Y_n)$. Moreover, a simple calculation and Lemma 8 yield

$$\begin{aligned} \text{Var}^*(V_{kijij'}) &\leq \mathbb{P}^*\{X_i \leq X_k^* \leq X_{i'}\} + \mathbb{P}^*\{Y_j \leq Y_k^* \leq Y_{j'}\} \\ &\leq \sup_s [\mathbb{F}_n(s + \delta_n) - \mathbb{F}_n(s)] + \sup_t [\mathbb{G}_n(t + \delta_n) - \mathbb{G}_n(t)] \\ &\leq 2\delta_n + 2n^{-1/2}\mathbb{M}_n(\delta_n) \\ &\leq K \max(\delta_n, M_n(\delta_n)/\sqrt{n}) \\ &\leq K \max(\delta_n, \sqrt{\delta_n \log n}/\sqrt{n}) \\ &= K\delta_n \end{aligned}$$

for n large enough, for almost all realizations (X_i, Y_i) and for some constant $K > 0$. Hence, by the union bound and Bernstein's exponential inequality for bounded random variables, we have, for some constant K_1 ,

$$\begin{aligned} &\mathbb{P}^* \left\{ \max_{i,i',j,j' \in \{1,\dots,n\}, |X_i - X_{i'}| < \delta_n, |Y_j - Y_{j'}| < \delta_n} |\alpha_n^*(X_i, Y_j) - \alpha_n^*(X_{i'}, Y_{j'})| > x \right\} \\ &\leq 2n^4 \exp(-K_1(\sqrt{n}x \wedge x^2\delta_n^{-1})), \end{aligned}$$

for all samples $(X_1, Y_1), \dots, (X_n, Y_n)$. By integrating the previous inequality over \mathbb{P} , we get the same inequality, but replacing \mathbb{P}^* by \mathbb{P} . Set $x = K_0\delta_n^{1/2}(\log n)^{1/2}$ and take the constant K_0 is sufficiently large, to obtain

$$\sum_{n=1}^{\infty} \mathbb{P} \{ \mathbb{M}_n^*(\delta_n) > K_0\delta_n^{1/2}(\log n)^{1/2} \} < +\infty.$$

Apply the Borel-Cantelli lemma to conclude the proof. \square

Analogous to the approximation of the process \mathbb{Z}_n by $\tilde{\mathbb{Z}}_n$ before, we introduce a simpler process $\tilde{\mathbb{Z}}_n^*$ to approximate \mathbb{Z}_n^* . Set

$$(34) \quad \tilde{\mathbb{Z}}_n^*(s, t) = \sqrt{n}(\mathbb{H}_n^* - \mathbb{H}_n)(s, t) - C_1(s, t)\sqrt{n}(\mathbb{F}_n^* - \mathbb{F}_n)(s) - C_2(s, t)\sqrt{n}(\mathbb{G}_n^* - \mathbb{G}_n)(t),$$

and

$$\tilde{\mathbb{Z}}_n^*(f) = \sum_{k=1}^{4L_n} \sigma_k \tilde{\mathbb{Z}}_n^*(s_k, t_k).$$

PROPOSITION 11. *Under conditions (C1) and (C2), we have*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[d_{BL_1}(\mathbb{Z}_n^*, \tilde{\mathbb{Z}}_n^*) \right] = 0.$$

Proof. First, we notice that, for any $\delta > 0$,

$$(35) \quad \mathbb{E} \left[d_{BL_1}(\mathbb{Z}_n^*, \tilde{\mathbb{Z}}_n^*) \right] \leq \delta + 2\mathbb{E} \left[\mathbb{P}^* \left\{ \sup_{f \in \mathcal{F}_n} \left| \mathbb{Z}_n^*(f) - \tilde{\mathbb{Z}}_n^*(f) \right| \geq \delta \right\} \right].$$

Hence it suffices to show that

$$(36) \quad \sup_{f \in \mathcal{F}_n} \left| \mathbb{Z}_n^*(f) - \tilde{\mathbb{Z}}_n^*(f) \right| \rightarrow 0$$

in \mathbb{P}^* -probability, for all realizations of $(X_1, Y_1), \dots, (X_n, Y_n)$ in an event Ω_n with $\mathbb{P}\{\Omega_n\} \rightarrow 1$. We now verify (36). Some straightforward adding and subtracting yields

$$\mathbb{Z}_n^*(s, t) = \bar{\mathbb{Z}}_n^*(s, t) + R_n^*(s, t)$$

with

$$\bar{\mathbb{Z}}_n^*(s, t) = \sqrt{n}\{\mathbb{H}_n^*(s, t) - \mathbb{H}_n(s, t)\} + \sqrt{n}\{C(\mathbb{F}_n s, \mathbb{G}_n t) - C(\mathbb{F}_n^* s, \mathbb{G}_n^* t)\}$$

and

$$\begin{aligned} R_n^*(s, t) &= \sqrt{n} \{C(\mathbb{F}_n^{*-} s, \mathbb{G}_n^{*-} t) - C(s, t)\} - \sqrt{n} \{C(s, t) - C(\mathbb{F}_n^* s, \mathbb{G}_n^* t)\} + \\ &\quad - \sqrt{n} \{C(\mathbb{F}_n s, \mathbb{G}_n t) - C(s, t)\} - \sqrt{n} \{C(\mathbb{F}_n^- s, \mathbb{G}_n^- t) - C(s, t)\} \\ &\quad + \{\alpha_n^*(\mathbb{F}_n^- s, \mathbb{G}_n^- t) - \alpha_n^*(s, t)\} \\ &\quad + \{\alpha_n(\mathbb{F}_n^{*-} s, \mathbb{G}_n^{*-} t) - \alpha_n(\mathbb{F}_n^- s, \mathbb{G}_n^- t)\} + \\ &\quad + \{\alpha_n^*(\mathbb{F}_n^{*-} s, \mathbb{G}_n^{*-} t) - \alpha_n^*(\mathbb{F}_n^- s, \mathbb{G}_n^- t)\} + \{\alpha_n^*(\mathbb{F}_n^- s, \mathbb{G}_n^* t) - \alpha_n^*(\mathbb{F}_n^- s, \mathbb{G}_n^- t)\}. \end{aligned}$$

Next, we observe that

$$\begin{aligned} d_{BL_1}(\mathbb{Z}_n^*, \tilde{\mathbb{Z}}_n^*) &\leq 4L_n \sup_{s,t} |\mathbb{Z}_n^*(s,t) - \tilde{\mathbb{Z}}_n^*(s,t)| \\ &\leq 4L_n \sup_{s,t} |\bar{\mathbb{Z}}_n^*(s,t) - \tilde{\mathbb{Z}}_n^*(s,t)| + 4L_n \sup_{s,t} |R_n^*(s,t)|. \end{aligned}$$

Now, use the same techniques as in the proof of Proposition 9, adapted to bootstrapped empirical processes. To be specific, with the aid of the following approximation results:

$$\sup_{0 < s < 1} |\alpha_n^*(s, 1)| / (s^{1/2}(1-s)^{1/2}) = O_{p^*}(1) \quad \text{almost surely,}$$

see Theorem 2.1, display (2.10) and Remark 2 at page 1457 in Csörgő and Mason (1989),

$$\sup_{s,t \in [0, \varepsilon_n] \cup [1-\varepsilon_n, 1]} [|\alpha_n^*(s, 1)| + |\alpha_n^*(1, t)|] = O(\varepsilon_n^{1/2}(\log n)^{1/2}) \quad \text{almost surely,}$$

see Lemma 10 above,

$$\sup_s |\beta_{n1}^*(s) + \alpha_n^*(s, 1)| = O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}) \quad \text{almost surely,}$$

see displays (2.10') and (2.12') in Theorem 2.1 of Csörgő and Mason (1989), we can show that $4L_n \sup_{s,t} |\bar{\mathbb{Z}}_n^*(s,t) - \tilde{\mathbb{Z}}_n^*(s,t)| + 4L_n \sup_{s,t} |R_n^*(s,t)| = O_{p^*}(L_n n^{-\xi})$ for some $\xi > 0$, in probability, proving the result.

For instance, to deal with R_n^* , note that

$$|\alpha_n^*(\mathbb{F}_n^- s, \mathbb{G}_n^- t) - \alpha_n^*(s, t)| \leq \mathbb{M}_n^*(\delta),$$

with $\delta = O(n^{-1/2}(\log \log n)^{1/2})$ almost surely. Consequently, by Lemma 10 this term is of order $O_{p^*}(n^{-a})$, for $0 < a < 1/4$. Another term in R_n^* can be handled as follows:

$$|\alpha_n(\mathbb{F}_n^{*-} s, \mathbb{G}_n^{*-} t) - \alpha_n(\mathbb{F}_n^- s, \mathbb{G}_n^- t)| \leq \mathbb{M}_n(\delta)$$

with $\delta = O_{p^*}(n^{-1/2})$, so that the term on the left is of order $O_{p^*}(n^{-1/4}(\log n)^{1/2})$ by Lemma 8. The other terms can be treated in the same way. \square

5.2. PROOF OF THEOREM 1. We need to show that $d_{BL_1}(\mathbb{Z}_n, \mathbb{Z}_n^*) \rightarrow 0$ in probability, as $n \rightarrow \infty$. By triangle inequality, we have, with probability one,

$$\begin{aligned} d_{BL_1}(\mathbb{Z}_n, \mathbb{Z}_n^*) &= \sup_{h \in BL_1} |\mathbb{E}[h(\mathbb{Z}_n)] - \mathbb{E}^*[h(\mathbb{Z}_n^*)]| \\ &\leq \sup_{h \in BL_1} |\mathbb{E}[h(\mathbb{Z}_n) - h(\tilde{\mathbb{Z}}_n)]| + \sup_{h \in BL_1} |\mathbb{E}[h(\tilde{\mathbb{Z}}_n)] - \mathbb{E}^*[h(\tilde{\mathbb{Z}}_n^*)]| \\ &\quad + \sup_{h \in BL_1} |\mathbb{E}^*[h(\tilde{\mathbb{Z}}_n^*) - h(\mathbb{Z}_n^*)]| \\ &= d_{BL_1}(\mathbb{Z}_n, \tilde{\mathbb{Z}}_n) + d_{BL_1}(\tilde{\mathbb{Z}}_n, \tilde{\mathbb{Z}}_n^*) + d_{BL_1}(\tilde{\mathbb{Z}}_n^*, \mathbb{Z}_n^*). \end{aligned}$$

Since Proposition 9 shows that $d_{BL_1}(\mathbb{Z}_n, \tilde{\mathbb{Z}}_n) \rightarrow 0$ as $n \rightarrow \infty$ and Proposition 11 implies that $d_{BL_1}(\tilde{\mathbb{Z}}_n^*, \mathbb{Z}_n^*) \rightarrow 0$, in probability, as $n \rightarrow \infty$, it remains to show that $d_{BL_1}(\tilde{\mathbb{Z}}_n, \tilde{\mathbb{Z}}_n^*) \rightarrow 0$, in probability, as $n \rightarrow \infty$. We recall that

$$\begin{aligned} \tilde{\mathbb{Z}}_n(f) &= \sum_{k=1}^{4L_n} \sigma_k \tilde{\mathbb{Z}}_n(s_k, t_k) \\ &= \sum_{k=1}^{4L_n} \sigma_k \int f_k(x, y) d\alpha_n(x, y), \end{aligned}$$

for

$$f_k(x, y) = 1\{x \leq s_k, y \leq t_k\} - C_1(s_k, t_k)1\{x \leq s_k\} - C_2(s_k, t_k)1\{y \leq t_k\}.$$

Now, let $h_f(x, y) = \sum_{k=1}^{4L_n} \sigma_k f_k(x, y)$ so that

$$(37) \quad \tilde{\mathbb{Z}}_n(f) = \int h_f d\alpha_n,$$

and we can derive in the same way

$$(38) \quad \tilde{\mathbb{Z}}_n^*(f) = \int h_f d\alpha_n^*.$$

We now apply Theorem 3 in Radulović (2012), stated as Theorem 12 in the appendix for convenience. We need to verify that

- the three classes

$$\begin{aligned} \mathcal{G}_k^a &= \{1\{x \leq s_k, y \leq t_k\}, 0 \leq s_k, t_k \leq 1\}, \\ \mathcal{G}_k^b &= \{C_1(s_k, t_k)1\{x \leq s_k\}, 0 \leq s_k, t_k \leq 1\}, \\ \mathcal{G}_k^c &= \{C_2(s_k, t_k)1\{y \leq t_k\}, 0 \leq s_k, t_k \leq 1\}, \end{aligned}$$

have VC-indices V_k^a , V_k^b and V_k^c , respectively, with $\sum_{k=1}^{4L_n} (V_k^a + V_k^b + V_k^c) \leq K(\log n)^\gamma$ for some finite constant K and some $0 < \gamma < 1$.

- the class $\mathcal{H}_n = \{h_f : f \in \mathcal{F}_n\}$ has an envelope $H(x, y)$ with $\mathbb{E}[H^4(X, Y)] < \infty$.

First we verify the VC property. The class \mathcal{G}_k^a is VC with VC-dimension $V_k^a = 3$ (Van der Vaart and Wellner (1996, page 135)), while the class \mathcal{G}_k^b is a subclass of the class of functions $d1\{a \leq x \leq b\}$ with $a, b \in \mathbb{R}$ and $d > 0$. This class has a VC index 3 : see van der Vaart and Wellner (1996, Problem 20, page 153). The same reasoning applies for the class \mathcal{G}_k^c , and consequently

$$\sum_{k=1}^{4L_n} (V_k^a + V_k^b + V_k^c) \leq 36L_n \leq K(\log n)^\gamma$$

for some $K < \infty$. It remains to verify the envelope condition. Writing

$$g_{x,y}(s, t) = 1\{x \leq s, y \leq t\} - C_1(s, t)1\{x \leq s\} - C_2(s, t)1\{y \leq t\}$$

we see that

$$h_f(x, y) = \sum_{k=1}^{L_n} c_k g_{x,y}(B_k)$$

for $c_k = \pm 1$ and $B_k = [s_k^1, s_k^2] \times [t_k^1, t_k^2]$. Here $\{s_k^1, t_k^1, s_k^2, t_k^2\}_{k=1}^{L_n}$ are the appropriately relabeled $\{s_k, t_k\}_{k=1}^{4L_n}$ and the operation $\phi(B_k)$, for any function $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$, is defined as

$$\phi(B_k) = \phi(s_k^2, t_k^2) - \phi(s_k^2, t_k^1) - \phi(s_k^1, t_k^2) + \phi(s_k^1, t_k^1).$$

Furthermore, writing

$$\begin{aligned} I_{x,y}(s, t) &= 1\{x \leq s, y \leq t\}, \\ II_{x,y}(s, t) &= C_1(s, t)1\{x \leq s\}, \\ III_{x,y}(s, t) &= C_2(s, t)1\{y \leq t\}, \end{aligned}$$

we have

$$\begin{aligned} |h_f(x, y)| &= \left| \sum_{k=1}^{L_n} c_k g_{x,y}(B_k) \right| \\ &\leq \sum_{k=1}^{L_n} |I_{x,y}(B_k)| + \sum_{k=1}^{L_n} |II_{x,y}(B_k)| + \sum_{k=1}^{L_n} |III_{x,y}(B_k)| \end{aligned}$$

Trivially, since rectangles are disjoint, the first term $\sum_{k=1}^{L_n} |I_{x,y}(B_k)|$ is bounded by 1. The estimate for the two remaining terms on the right are identical (just use C_2 instead of C_1), so we bound $\sum_{k=1}^{L_n} |II_{x,y}(B_k)|$ only. We first observe that, after a little algebra,

$$II_{x,y}(B_k) = [C_1(s_k^2, t_k^2) - C_1(s_k^2, t_k^1)] 1\{s_k^1 < x \leq s_k^2\} + 1\{x \leq s_k^1\} C_1(B_k).$$

This identity yields

$$\begin{aligned} \sum_{k=1}^{L_n} |II_{x,y}(B_k)| &\leq \sum_{k=1}^{L_n} |C_1(s_k^2, t_k^2) - C_1(s_k^2, t_k^1)| 1\{s_k^1 < x \leq s_k^2\} + \sum_{k=1}^{L_n} |1\{x \leq s_k^1\} C_1(B_k)| \\ &\leq 2 \sum_{k=1}^{L_n} 1\{s_k^1 < x \leq s_k^2\} + \sum_{k=1}^{L_n} |C_1(B_k)| \\ &\leq 2 + TV(C_1). \end{aligned}$$

The last inequality follows from $\|C_1\|_\infty \leq 1$ (since copulas are Lipschitz with Lipschitz constant 1) and our assumption (C1) that $TV(C_1)$, the total variation of C_1 , is finite. Consequently, the class \mathcal{H}_n has envelope $5 + TV(C_1) + TV(C_2)$.

We can now apply Theorem 12 to conclude that $d_{BL_1}(\tilde{Z}_n, \tilde{Z}_n^*) \rightarrow 0$, in probability, as $n \rightarrow \infty$. \square

5.3. PROOF OF THEOREM 3. We proceed as in the proof of Theorem 1. We write $\hat{C} = C_{\hat{\theta}}$ and $\hat{C}^* = C_{\hat{\theta}^*}$. Recall that

$$\mathbb{Y}_n = \mathbb{Z}_n - \sqrt{n}(\hat{C} - C).$$

We may replace \mathbb{Z}_n by $\tilde{\mathbb{Z}}_n$ with impunity since

$$\begin{aligned}
& d_{BL_1}(\mathbb{Y}_n, \tilde{\mathbb{Z}}_n - \sqrt{n}(\hat{C} - C)) \\
& \leq \delta + 2\mathbb{P} \left\{ \sup_{f \in \mathcal{F}_n} \left| \mathbb{Y}_n(f) - \tilde{\mathbb{Z}}_n(f) + \sqrt{n}(\hat{C} - C)(f) \right| \geq \delta \right\} \\
& = \delta + 2\mathbb{P} \left\{ \sup_{f \in \mathcal{F}_n} \left| \mathbb{Z}_n(f) - \tilde{\mathbb{Z}}_n(f) \right| \geq \delta \right\} \\
& \rightarrow \delta \text{ as } n \rightarrow \infty,
\end{aligned}$$

for every $\delta > 0$, as in the proof of Proposition 9. Next, by the mean value theorem and assumptions (C3) and (C4), we have

$$\begin{aligned}
\sqrt{n}(\hat{C} - C)(s, t) &= \sqrt{n}(\hat{\theta} - \theta_0)' \dot{C}_{\theta_0}(s, t) + \sqrt{n}(\hat{\theta} - \theta_0)' \{ \dot{C}_{\tilde{\theta}}(s, t) - \dot{C}_{\theta_0}(s, t) \} \\
&\quad \text{for some } \tilde{\theta} \text{ between } \hat{\theta} \text{ and } \theta_0 \\
&= \left(\int \psi d\alpha_n + n^{1/2} \varepsilon_n \right)' \dot{C}_{\theta_0}(s, t) + \sqrt{n}(\hat{\theta} - \theta_0)' \{ \dot{C}_{\tilde{\theta}}(s, t) - \dot{C}_{\theta_0}(s, t) \} \\
&= \left(\int \psi d\alpha_n \right)' \dot{C}_{\theta_0}(s, t) + R_n(s, t)
\end{aligned}$$

for some remainder term R_n that satisfies

$$\begin{aligned}
|R_n(s, t)| &\leq n^{1/2} \|\varepsilon_n\| \|\dot{C}_{\theta_0}(s, t)\| + K n^{1/2} \|\hat{\theta} - \theta_0\|^{1+\nu} \\
&= O_p(n^{1/2} \|\varepsilon_n\| + n^{-\nu/2}) \\
&= o_p(1/L_n).
\end{aligned}$$

This bound holds uniformly in s and t . Consequently, for

$$\tilde{\mathbb{Y}}_n(f) = \sum_{k=1}^{4L} \sigma_k \tilde{\mathbb{Y}}_n(s_k, t_k)$$

based on

$$\tilde{\mathbb{Y}}_n(s, t) = \tilde{\mathbb{Z}}_n(s, t) - \left(\int \psi d\alpha_n \right)' \dot{C}_{\theta_0}(s, t),$$

we have

$$d_{BL_1}(\tilde{\mathbb{Z}}_n - \sqrt{n}(\hat{C} - C), \tilde{\mathbb{Y}}_n) = d_{BL_1}(\tilde{\mathbb{Y}}_n - R_n, \tilde{\mathbb{Y}}_n).$$

Since

$$\sup_f |R_n(f)| \leq L_n \sup_{s,t} |R_n(s, t)| \rightarrow 0$$

in probability, we get $d_{BL_1}(\tilde{\mathbb{Z}}_n - \sqrt{n}(\hat{C} - C), \tilde{\mathbb{Y}}_n) \rightarrow 0$, as $n \rightarrow \infty$. We conclude that

$$\limsup_{n \rightarrow \infty} d_{BL_1}(\mathbb{Y}_n, \tilde{\mathbb{Y}}_n) = 0.$$

For the bootstrap counterpart, we can argue in the same way. Using the expansion

$$\sqrt{n}(\hat{C}^* - \hat{C})(s, t) = \left(\int \psi d\alpha_n^* \right)' \dot{C}_{\theta_0}(s, t) + R_n^*(s, t)$$

for some remainder term R_n^* that satisfies

$$\sup_{s,t} |R_n^*(s, t)| \leq K_0 n^{1/2} \|\varepsilon_n^*\| + K_1 n^{1/2} \|\hat{\theta} - \theta_0\|^{1+\nu} + K_2 n^{1/2} \|\hat{\theta}^* - \hat{\theta}\|^{1+\nu},$$

for some finite constants K_0, K_1 and K_2 . We check that the processes \mathbb{Y}_n^* and $\tilde{\mathbb{Y}}_n^*$ are close, in the sense that

$$d_{BL_1}(\mathbb{Y}_n^*, \tilde{\mathbb{Y}}_n^*) \rightarrow 0$$

in probability, as $n \rightarrow \infty$. Here, $\tilde{\mathbb{Y}}_n^*$ is based on

$$\tilde{\mathbb{Y}}_n^*(s, t) = \tilde{\mathbb{Z}}_n^*(s, t) - \left(\int \psi d\alpha_n^* \right)' \dot{C}_{\theta_0}(s, t).$$

Note that $\tilde{\mathbb{Y}}_n(f) = \sum_k \sigma_k \tilde{\mathbb{Y}}_n(s_k, t_k) = \int (\sum_k \sigma_k g_k) d\alpha_n$ with

$$\begin{aligned} g_k(x, y) &= 1\{x \leq s_k, y \leq t_k\} - C_1(s_k, t_k)1\{x \leq s_k\} - C_2(s_k, t_k)1\{y \leq t_k\} \\ &\quad - (\psi(x, y))' \dot{C}_{\theta_0}(s_k, t_k). \end{aligned}$$

As in the proof of Theorem 1, it remains to verify the two conditions of Theorem 12. Since the only difference with the proof of Theorem 1 is the addition of the term $(\psi(x, y))' \dot{C}_{\theta_0}(s_k, t_k)$, we concentrate on the class of functions $(\psi(x, y))' \dot{C}_{\theta_0}(s_k, t_k)$. Since it is a subclass of $c' \psi(x, y)$ with $c \in \mathbb{R}^p$, its VC dimension trivially is equal to p . Moreover, it is not hard to see from the proof of Theorem 1 that

$$\left| \sum_{k=1}^{4L_n} \sigma_k g_k(x, y) \right| \leq 5 + TV(C_1) + TV(C_2) + \|\psi(x, y)\| TV(\dot{C}_{\theta_0}),$$

Since $\mathbb{E}[\|\psi(X, Y)\|^4] < \infty$, the conditions of Theorem 12 are met, and we conclude that $d_{BL_1}(\tilde{\mathbb{Y}}_n, \tilde{\mathbb{Y}}_n^*) \rightarrow 0$ in probability, as $n \rightarrow \infty$. \square

5.4. PROOF OF PROPOSITION 5. From the proofs of Proposition 9 and Proposition 11, we see that

$$\sup_{(u,v) \in [0,1]^2} |\mathbb{Z}_n(u,v) - \tilde{\mathbb{Z}}_n(u,v)| = O_p(n^{-\mu}), \text{ and}$$

$$\sup_{(u,v) \in [0,1]^2} |\mathbb{Z}_n^*(u,v) - \tilde{\mathbb{Z}}_n^*(u,v)| = O_{p^*}(n^{-\mu})$$

almost surely, for some $\mu > 0$. The result follows after integration by parts. \square

5.5. PROOF OF COROLLARY 6. By the delta-method, $\{\tilde{\mathbb{Y}}_n(s,t), 0 \leq s, t \leq 1\}$ converges towards a Gaussian process in $\ell^\infty([0,1]^2)$. The proof of Theorem 3 shows that $\limsup_{n \rightarrow \infty} d_{BL_1}(\mathbb{Y}_n, \tilde{\mathbb{Y}}_n) = 0$. Hence, the process \mathbb{Y}_n converges weakly to the same limit as $\tilde{\mathbb{Y}}_n$. This proves the first claim. The second part of the Corollary is a straightforward consequence of Theorem 3 and the triangle inequality applied to the distance d_{BL_1} . \square

APPENDIX A.

Let X_1, \dots, X_n be independent random variables with probability measure P . Let \mathbb{P}_n be the empirical probability measure, putting mass $1/n$ at each observation, and let \mathbb{P}_n^* be the nonparametric bootstrap measure based on n independent observations from \mathbb{P}_n . We index the empirical process $\sqrt{n}(\mathbb{P}_n - P)$ and its bootstrap counterpart $\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ by functions f that belong to a sequence of classes \mathcal{F}_n .

THEOREM 12. *Let d_n be an integer sequence and, for each $1 \leq i \leq d_n$, let $\mathcal{G}_{i,n}$ be a VC class of functions with VC index $V_{i,n}$ and*

$$\sum_{i=1}^{d_n} V_{i,n} \leq K(\log n)^\gamma,$$

for some $K < \infty$ and $0 < \gamma < 1$. Set

$$\mathcal{F}_n = \left\{ f = \sum_{i=1}^{d_n} g_i : g_i \in \mathcal{G}_{i,n} \right\},$$

and suppose that there exists an envelope function $F \geq \sup_{f \in \mathcal{F}_n} |f|$, independent of n , with $\mathbb{E}[F^4(X)] < \infty$. Then,

$$d_{BL_1}(\{\sqrt{n}(\mathbb{P}_n - P)(f), f \in \mathcal{F}_n\}, \{\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)(f), f \in \mathcal{F}_n\}) \rightarrow 0,$$

in probability, as $n \rightarrow \infty$.

Proof. See Theorem 3 in Radulović (2012). \square

APPENDIX B.

Set $\mathbb{M}_n(\delta)$ as in (31) for $\delta \geq 0$, and define

$$\psi(x) = 2x^{-2}\{(1+x)\log(1+x) - x\}, \quad x \in (-1, 0) \cup (0, \infty)$$

and $\psi(-1) = 2$ and $\psi(0) = 1$. This function is continuous and decreasing.

PROPOSITION 13. *There exist constants K_1 and K_2 such that*

$$(39) \quad \mathbb{P}\{\mathbb{M}_n(a) \geq \lambda\} \leq \frac{K_1}{a} \exp\left\{-\frac{K_2\lambda^2}{a}\psi\left(\frac{\lambda}{\sqrt{na}}\right)\right\}$$

for all $a \in (0, 1/2]$ and all $\lambda \in [0, \infty)$.

Proof. See Proposition A.1 of Segers (2012). Segers credits Einmahl (1987) who proved this bound for the independence copula $C(u, v) = uv$ and Tsukahara (private communication) for observing that the inequality holds for any copula. \square

REFERENCES

1. M.J. Appel, R. Labarre, and D. Radulović (2004). On Accelerated Random Search, *SIAM Journal on Optimization*, **14**(3), 708-731.
2. D. Berg (2009). Copula goodness-of-fit testing: An overview and power comparison. *European Journal of Finance*, **15**, 675 – 701
3. Bickel, P.J. and Freedman, D.A. (1983). Bootstrapping Regression Models with Many Parameters. In *A Festschrift for Erich L. Lehmann*, Editors P. J. Bickel, K. Doksum, J. L. Hodges. Wadsworth Statistics/Probability Series.
4. A. Bücher and H. Dette (2010). A note on bootstrap approximations for the empirical copula process. *Statistics and Probability Letters*, **80**, 1925-1932.
5. S. Csörgő and D.M. Mason (1989). Bootstrapping Empirical Functions. *Annals of Statistics*, **17**(4), 1447-1471.

6. P. Deheuvels (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Académie Royale de Belgique. Bulletin de la Classe des Sciences / Koninklijke Belgische Academie. Mededelingen van de Klasse der Wetenschappen*, **5**(65), 274 – 292.
7. J.H.J. Einmahl (1987). *Multivariate empirical processes*. CWI-tract 32. Centrum voor Wiskunde en Informatica: Amsterdam.
8. J.-D. Fermanian (2005). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis*, **95**(1) 119 – 152.
9. J.-D. Fermanian, D. Radulović and M. H. Wegkamp (2004). Weak convergence of empirical copula processes. *Bernoulli*, **10**, 847 – 860.
10. J.-D. Fermanian (2012). An overview of the goodness-of-fit test problem for copulas. To appear.
11. J.-D. Fermanian and M. H. Wegkamp (2012). Time-dependent copulas. *Journal of Multivariate Analysis*, **110**, 19 – 29.
12. P. Gänslér and W. Stute (1987). *Seminar on Empirical Processes*. Birkhäuser Verlag, Basel, Switzerland.
13. C. Genest, K. Ghouli and L.-P. Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543 – 552.
14. C. Genest and B. Remillard (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré (Probabilités et Statistiques)*, **44**(6), 1096 – 1127.
15. C. Genest, B. Remillard & D. Beaudoin (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, **44**, 199-213.
16. W. Hoeffding (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, **19**, 293-325.
17. D. Mason (1981). Bounds for Weighted Empirical Distribution Functions. *Annals of Probability*, **9**, 881 – 884.
18. D. Radulović (1998). Can we bootstrap even if CLT fails? *Journal of Theoretical Probability*, **11**, 813-830.
19. D. Radulović (2012). Direct Bootstrapping Technique and its Application to a Novel Goodness of Fit Test. *Journal of Multivariate Analysis*, **107**, 181-199.
20. B. Remillard and O. Scaillet (2009). Testing for equality between two copulas *Journal of Multivariate Analysis*, **100**, 377 – 386.
21. O. Scaillet (2007). Kernel-based goodness-of-fit tests with fixed design smoothing parameters. *Journal of Multivariate Analysis*, **98**, 533 – 543.
22. J. Segers (2012). Asymptotic of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, **18**, 764-782.
23. J. H. Shih and T. A. Louis (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, **51**, 1384 – 1399.
24. G.R. Shorack and J. A. Wellner (2009). *Empirical Processes with Applications to Statistics*, 2nd Edition, SIAM.
25. A.W. van der Vaart (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge.

JEAN-DAVID FERMANIAN. CREST-ENSAE, MALAKOFF, FRANCE.

DRAGAN RADULOVIĆ. DEPARTMENT OF MATHEMATICS, FLORIDA ATLANTIC UNIVERSITY.

MARTEN H. WEGKAMP. DEPARTMENT OF MATHEMATICS & DEPARTMENT OF STATISTICAL SCIENCE, CORNELL UNIVERSITY.